

图书馆情报学系列教程

TUSHUGUANXUE QINGBAOXUE

XILIE JIAOCHENG

文献计量学教程

王崇德 编著



G350
21

来新夏 主编

图书馆学情报学系列教程
TUSHUGUANXUE QINGBAOXUE XILIE

JIAOCHENG

文献计量学教程

王崇德 编著

南开大学出版社

内 容 简 介

文献计量学是当今科学定量化趋势中发展起来的新学科。全书从文献作者分布、文献分散、词频分布、文献增长、文献老化以及引文分析等方面，论述其数学和统计学规律。给以必要的数学预备知识，又紧密环绕文献学、情报学的科学计量予以阐述，举例论证。可做情报学、档案学、图书馆学和科学社会学等系科的教材，亦可供科学学、未来学、情报学、决策科学工作者参考。

文献计量学教程

王崇德 编著

南开大学出版社出版

(天津八里台南开大学校内)

邮政编码:300071 电话:34.9318

新华书店天津发行所发行

河北省邮电印刷厂印刷

1990年6月第1版

1990年6月第1次印刷

开本: 850×1168 1/32

印张: 10.625 插页: 2

字数: 285千

印数: 1—1,500

ISBN7-310-00192-3/G·20

定价: 2.85元

序 言

王崇德同志是我国情报界知名学者，他的学术造诣是很深的，已为情报学界所公认。特别在文献计量学方面，他已达到我国一流水平。

他撰著的《文献计量学教程》很有特色，内容丰富，观点新颖而正确。文献计量学是要以数学作工具来描述的，但作者不唯数学。在内容上文理兼容，反映了当代自然科学与社会科学交叉发展的总趋势。

文献计量学，是科学学、未来学、情报学、决策科学工作者都需要的，读者也是比较多的。

王崇德同志所著《文献计量学教程》，适合作情报学、档案学、图书馆学、科学社会学等各系学生教材使用，也可作教师教学参考。教师、学生都可选读。

欣闻南开大学出版社出版此书，是非常必要的。我在情报界从事情报学研究27年，深感此种知识普及的必要，颇知出版此书的社会需要，不揣冒昧向全国图书情报界推荐。

中国科学技术培训中心主任

教授 杨沛霆

一九八七·九·十于北京

目 录

序言	杨沛霆
第一章 绪论	(1)
第一节 文献计量学的产生与发展	(1)
第二节 文献计量学的研究对象和特点	(9)
第三节 文献计量学的研究方法	(14)
第二章 文献计量学的测量	(25)
第一节 定性测量和定量测量	(25)
第二节 测量级别	(26)
第三节 测量的有效性和可靠性	(32)
第四节 测量的误差	(36)
第三章 统计学基础知识	(38)
第一节 简单的概率知识	(38)
第二节 随机变量及其分布	(43)
第三节 正态分布与中心极限定理	(57)
第四节 抽样与抽样分布	(60)
第五节 统计推断	(67)
第六节 回归分析	(72)
第四章 文献作者分布规律	(84)
第一节 洛特卡定律	(84)
第二节 洛特卡以后的文献作者分布规律	(94)
第三节 文献作者分布规律的应用	(100)
第五章 文献分散规律	(113)
第一节 布拉德福定律	(113)
第二节 布拉德福定律的发展	(118)
第三节 布拉德福定律的歧异性与一致性	(134)

第四节	布拉德福定律的理论基础	(144)
第五节	布拉德福定律的应用	(146)
第六章	词频分布规律	(166)
第一节	齐夫定律的开发过程	(166)
第二节	齐夫定律的概况	(170)
第三节	中文词频分布规律	(177)
第四节	齐夫定律的理论基础	(181)
第五节	齐夫定律的应用	(190)
第七章	文献增长规律	(195)
第一节	指数增长规律	(195)
第二节	滑动指数模型与直线模型	(204)
第三节	逻辑曲线增长模型	(211)
第四节	综合性模型	(214)
第五节	文献增长的原因及其规律的应用	(222)
第八章	文献老化规律	(229)
第一节	科技文献的老化	(230)
第二节	文献老化的数学模型	(236)
第三节	文献老化规律的应用	(247)
第九章	引文分析	(253)
第一节	引文分析及其特点	(253)
第二节	引文分析的指标与来源资料	(260)
第三节	引文分析的方法	(268)
第四节	引文分析的具体应用	(288)
第十章	文献计量学各定律之间的相关性	(307)
第一节	布一齐一洛定律的一致性	(307)
第二节	文献增长与老化的关系	(320)
第三节	文献分散与老化的关系	(321)

后 记

第一章 绪 论

现代科学技术有一个显著的特点，就是科学知识的数学化。在自然科学各学科竞相精密定量化的同时，自20世纪以来，一个源于希腊词“μετρον”（测量）的英语词素“metrice”（计量学），也在社会科学和思维科学的许多领域中频繁地出现，形成许多重要的定量化学科分支。如，econometrics（经济计量学）、sciometrics（社会计量学）、neutrometrics（神经计量学）。其中文献计量学（bibliometrics）也应运逐渐孕成。一篇以被引用作者来研究情报学智力结构的权威文章，令人信服地发现文献计量学的作者群，如布鲁克斯（Brookes）、莱恩（Line）、高夫曼（Goffman）、布拉德福（Bradford）、桑迪森（Sandison）等人居于核心〔1〕。据该文作者的解释是：上述作者的文章经常而大量地被情报交流和情报检索两方面论文作者所引用，意味着文献计量学日益趋向情报学的中心。

第一节 文献计量学的产生与发展

俄国著名思想家普列汉诺夫曾讲过：人类社会的科学能够而且也应当变为与自然科学同样严密的科学。他的这一主张，八十多年来一一应验了。促进与推动科学定量化的因素既有外在的——社会的，也有内在的——学科本身的。

一、外在因素

有以下的外在因素影响具体的学科，使它们无一例外地走上定量化的道路〔2〕。

1. 构造理论的需要

理论的优势对任何一门学科来说，都已成为至理明断。著名的瑞典化学家阿累尼乌斯就曾说过：“理论是科学知识领域中最重要推动力。”

任何一门学科的最高境界是构造科学理论，这时，定量化的概念性语言是不可或缺的。

按照现代科学方法论的观点，构造科学理论大致是这样进行的：首先建立模型，以期把握、逼近研究对象的原型，直接反映经验层次或科学事实。作为抽象形式的模型有很大程度的直觉和形象的色彩，然后用概念性语言通过描述模型来揭示对象的本质属性和相对关系。这时构造理论的任务可以说是完成了。

但是，如果再继之运用数学语言描述，那无异于画龙点睛而妙笔生花，使理论的叙述大为增色。因为这样可给予理论量的规定性，把模型的定性概念变换为数学形式，使模型和理论符号化、形式化，这样便可以大幅度地提高理论的精度与力度。

1948年申农(C.E.Shannon)等人发表了《通讯的数学原理》，就是通过定量化的途径，才导致信息论的诞生。尽人皆知，申农的成就是在哈特莱(E.Hartley)的科研成果基础上获得的。1928年哈特莱提出信息是除开内容意义之外的符号或者文字的排列。例如在K种符号的符号系列中，将N个符号排成序列，便有 K^N 种方式。这个序列所接受的信息量可以表示为： $H=N\log K$ 。但这个公式只有描述的意义，却没有实际量化计算的价值。

申农考虑了噪音的影响，引进了概率的思想，从而提出了信息量的如下计算公式：

$$I = \log_2 \frac{P_1}{P_2}$$

式中的 P_1 和 P_2 分别为后验概率和先验概率。

从哈特莱到申农前后二十年，只有在信息量可以计算之后，信息论才最终取得了学术上的立足点。可见量化是发展现代科学，

构造先进的科学理论最有力的武器。

定量化的突出地位和决定性的作用早已毋庸置疑，已成为一种不可逆转的科学惯性力。包括有文献规律在内的情报学当然也概莫能外。定量化的主潮冲击着情报学，在外部为情报学的量化带来了压力和动力，推动文献计量学的产生与发展。

2. 区分现代信息的需要

进入20世纪以来，信息量剧增，以致人类社会呈现大量的信息堆积。区分、鉴别如此繁多的信息，靠定性的描述是难以企及的，也可以说是根本无法办到的。例如动植物种属之间的区分，传统的分类法仅能依靠形态上的种种“表观”信息来区分，但是往往既不彻底，也不分明。现代分子生物学以细胞色素来进行定量的比较，对亲缘关系的远近才有可能作出科学的判断。如黑猩猩与人的差异为零，而猴子为1，兔子为9，龟为15，蝇为25，小麦为35等，可见量化在区分、辨别信息方面是无与伦比的。目前，这已是大势所趋，情报学也不能不在这一浪潮中，从不自觉到自觉地走上量化的发展道路。

3. 方法论的需要

当代人比历史上任何时期都注重方法论。许多先进的成果由于采用先进的方法论而玉成。它们散发出诱人的榜样力量。采用先进的研究方法已成人心所向的愿望。40年代出现的“老三论”——系统论、控制论、信息论；70年代诞生的“新三论”——耗散结构理论、协同论和突变论，可以说是现代科学方法论宝库中最先进的武器。它们超学科、超领域地适用于包括情报学在内的一切学科，因而被人们誉为“横断方法”。“新三论”和“老三论”都有一个明显的特征，就是量化。新、老三论的广泛普及，使一切学科都表现出对量化空前的亲和力，这种炽烈的氛围也不能不影响着情报学。

恐怕还不止这三点。但以这三点为代表的趋势，就足以形成了一股强劲的现代科学潮流，可以把任何一个学科都卷进量化的洪

流中。

二、内在因素

情报学有哪些内在的因素，促使它迅速定量化呢？

1. 经验与科学的矛盾

情报学的起源是众说纷纭的，情报学的客观存在今天却是个事实。但是给人的感觉是，这个学科理论困乏，定性的描述偏多，凭经验来办事，靠直觉来模仿，还停留在经验层次上，距体系严谨、理论完善、知识丰富的情报科学，还有较大的距离。把经验转化为科学，这是情报学的当务之急。如何从经验层次上升到科学高度，摆脱经验局限和主观臆断，所依靠的手段之一就是定量化。“情报”定义的莫衷一是，很能说明这种现状。众多难以数计的定义，绝大多数是从经验出发的。情报定义的最后突破，极有可能像当年申农那样，以提出某种定量化公式而见分晓。因此为了迅速解决经验与科学的矛盾，情报学本身也迫切需要运用定量化来体现。

2. 应用与理论的矛盾

尽管情报学理论的大厦还不完善，但是它的应用却是既普遍而又有效的。情报产业日益发展，成为社会上一个独立的行业。情报学这种应用超前、理论滞后的矛盾状况，愈来愈严重。很明显，这种局面是不会长远继续下去的，否则一定会因为理论的落后，终究要限制应用的普及。改变与改善这种局面的唯一途径，不是扼制应用的普及与进步，而应迅速提高理论水平，使两者能相互辉映和促进。这里定量化就是改变现状的最好选择。能够也只能通过定量化的途径，使理论得到充实和发展。

3. 封闭与开放的矛盾

情报学向以开放著称，它能大量吸收相关科学的先进的理论与方法，用以充实与完善自己。可是任何一个独立的学科是要追求自己的特殊性的，能表现出独到的个性，乃是这一学科成熟的重要标志。兼容并蓄式的开放，往往要干扰乃至破坏情报学的个性。因此封闭与开放之间势必要建立一种和谐才能恰到好处。这种局面的建

立就要靠定量化，用定量化来判断那些援引、殷鉴的自外而来的观点与方法。这种外来的观点与方法要情报学化，那也应是经过定量处理和再现，使之自然而然地进入情报学框架的。

综上所述，有充足理由论证情报学定量化是必要的，也是必需的。文献计量学的发展过程也反映出这些因素的巨大支配作用。

三、文献计量学的沿革与历程

文献计量学 (bibliometrics) 这一术语是1969年由英国学者普里查德 (A.Pritchard) 首先提出来的〔8〕。现在较为一致的意见是，文献计量学并非一个独立的学科，仅仅是数学方法与统计学方法的综合〔4〕。

人类对文献施以定量化加工处理，可以回溯到1911年。苏联情报学家们争辩说，文献计量学的历史是俄国化学家、院士瓦尔金 (П.Валгин) 开创的。他于1911年首次采用分析引文的方法，研究包括俄国在内的一些国家的化学家对化学所做的贡献。

为此，他搜集并研究了当时化学史上最著名的著作，其中包括俄国化学家H·A·孟拿特金的《化学观念发展概要》(1888年)，德国历史学家E·B·迈文的《化学史》(1905年)，德国历史学家A·拉登堡的《化学发展史报告》(1905年)，德国历史学家L·达尔姆泰施特尔的《自然科与技术史手册》(1908年)，英国化学史学家A·吉尔迪森的《化学简史》(1911年)和瓦尔金本人的《从历史的连续性来看解答理论》(1910年)。

瓦尔金依据上述图书所注有的书目，通过计算某个国家学者的著作引文来确定该国在发展化学方面的贡献。瓦尔金采用的计量方法应当说是很有说服力的。

1917年文献学家科尔 (F.T.Cole) 和伊尔斯 (N.B.Eales) 合作，在《科学进展》刊物上著文，介绍了现在仍然沿用的最基本的书目计量技术——对出版物按出版国和学科来计量与分类。他们将自1543年至1860年之间有关比较解剖学的出版物进行了分析，也就是对图书和期刊文章的有关题目进行计数，并按国家加以分组。

结果不但系统地查明了比较解剖学家在这一时期的研究重点及发展趋势，而且也掌握了这一时期文献在世界各国的分布。如果仍然沿用定性分析的话，那是达不到这样的水平的。

据G·R·威汀考证，1923年，英国图书馆学家霍姆(E.W. Hulme)在科尔等人工作基础上提出“书目统计学”〔5〕。

早在1948年，英国专业图书馆和情报机构协会的会议上，著名图书馆学家阮岗纳赞(S.R. Ranganathan)就呼吁：“图书馆家们必须仿效生物学计量、经济学计量和心理学计量来发展图书馆学计量”，因为“图书馆工作与服务涉及大量的数字内容”〔6〕。

1969年，普里查德正式采用了文献计量学这一术语。

时至今日，文献计量学的学术地位已经确立。有关的争议只是局部性的，而不是方向性的。

英国以普里查德为首的学者的观点是：文献计量学是在统计书目学基础上产生的，而且作为后者的完善与发展，应视为目录学的分支学科。

苏联的一些学者认为，文献计量学与科学计量学的区分，是不必要也是不可能的。两者均系用各种定量方法研究科学发展过程和规律的新学科。其代表人物是哈通(С.Д. Хаѣтун)，他在其代表作《科学计量学》中提出的科学计量学方法与文献计量学方法，大多是交叉的。它们是：引文分析，文摘杂志分析，一次文献的统计分析，科学文献中专门术语的分析，科学发展拨款进程的分析，科学研究机构和科学研究人员等数量的定量分析。

在1976年美国形成的概念是：“文献计量学是情报科学的特殊方法的综合。”为证实美国的这一“方法综合”的概念，沃维连涅(O. Wauvilny)做了详尽的调查。他选用1911—1982年发表的使用文献计量方法的著作1000件，其中运用引文分析的就多达684件(占68.4%)。各学科领域采用这种方法进行研究的情况如表1-1所示。

从这684篇出版物的分析可以看出，这些学术著作的出版年代

表 1-1 各学科利用引文分析法的情况

序号	学 科	主 要 刊 物	文章数	百分比
1	情报学	《科技情报》第1、2辑,《美国情报会志》,《美国文献》,《情报处理与管理》,《情报存储与检索》,《情报学》,《联机评论》,《近期目次》等	258	37.8
2	图书馆学	《苏联科技图书馆》,《图书馆资料与技术服务》,《图书馆季刊》,《图书馆学与文献年刊》,《图书馆学业务总览》	81	11.8
3	科学学	《科学学与情报学》,《科学计量学》	60	8.7
4	社会学	《科学的社会研究》,《社会教育》,《美国社会学评论》	46	6.9
5	化 学	《工厂实验室》,《化学教育杂志》,《化学文献杂志》	12	1.6
6	心理学	《美国心理学杂志》	6	0.9
7	物理学、 经济学、 冶金学、 控制论、 医 学、 气象学等	《苏联科学院通报》,《科学与生活》,《科学》,《科学月刊》,《科学进展》,《自然》,《美国电气与电子工程师学会工程管理汇刊》,《气象学新闻》,《计算机协会杂志》,《经济文献杂志》	221	32.3

可以大致划分为三个时期:

1. 1911——1947年为图书馆学时期,此时引文分析在图书馆学学术著作中用得最多,主要是用以揭示某些科学领域最急需的定期刊物。

2. 1948——1977年为情报学时期,这一阶段里引文分析在情报学学术著作中用得最多。主要用来揭示学者与专家们对情报的需求,以确定出版物的老化情况;揭示有关学科中被认为最有用的出版

物名单；预测某些学术流派或者科学领域的发展情况；对科学文献进行定量或定性测定，评价情报检索和情报服务等效率。

3. 1978——1982年为情报学—科学学时期，这一期间引文分析作为解决情报学问题的一种重要方法，又常常活跃在科学学研究领域之中，尤其是在国际期刊《科学计量学》问世之后。

无论如何，文献计量学方法已成为普遍而有效的研究方法，那么它作为一门分支学科也应当不成问题。为了加速文献计量学的普及进程，1980年国际文献联合会（FID）设立了情报计量学委员会（Committee on Informetrics, FID/IM）（文献计量学与情报学在内涵与外延上是有所不同的，但彻底的区分是很困难的），并且制定一个近期计划。要点是：

1. 情报计量学的设置准备（1981——1983年）

要解决并提供情报计量学教育、训练的基本设备与用具；提高情报人员在专业文献和情报学本身研究中使用数学手段的意识。

2. 建立情报计量学教育的方法（1984年—— ）

编制培养情报学的专家的情报计量学方法与教材。

3. 情报计量学设置课程的评价（1985年—— ）

将情报计量学委员会的课程设置概念加以具体化，以便能够在实际中使用。

4. 情报计量学的术语汇编（1982——1986年）

通过编制术语汇编，提高情报计量学中比较、鉴定各种成果的可能性。

该委员会的工作虽然中断过一段时间，但很快就恢复了活动。1982年4月在捷克斯洛伐克召开会议，研究了文献计量学的课程设置。其课程分为两部分：基础必修课及选修课。必修课为集合论、布尔代数、逻辑学、图论、组合论、概率与统计、行列式与矩阵等。

综上所述，文献计量学有其内外发展的动力与压力，加之国际学术组织的协调，其发展速度与规模必将是非常迅速和广泛的。它在图书情报学理论建设，实际应用，科学管理等方面必将日见鼎力

之功。

第二节 文献计量学的研究对象和特点

许多人都曾为文献计量学下过定义，以求对这一领域的研究对象和特点加以概括。主要的有代表性的观点如下：

1. “通过对书面交流的统计并对其方方面面的分析，以看出书面交流的过程，及某个专业的性质和发展的方向。”（休姆，1923年）

2. “把数学和统计学用于图书和其他文字通讯载体的科学。”（普里查德，1969年）

3. “应用数学和统计学方法，通过计算和分析书面交流的各个方面，来更清楚地表达书面交流过程和科学发展的特点和进程。”（兰沃尼，1969年）

4. “有关被记录下的论文以对其行为特征进行定量处理的科学。”（R·A·范索尼，1969年）

5. “通过文献、图书馆工作者、用户行为和特性的定量分析，实现图书馆和情报中心的情报加工和情报处理的科学。”（拉维昌德拉·劳，1972年）

6. “以文献数量为基础，论述与预测科学技术规律与现象的学问。”（王崇德，1980年〔7〕）

尽管说法各异，但是记录下来的知识——文献却是无一回避，所以文献及围绕文献有关的各种变量应当是其研究对象。

一、研究对象

（一）与文献相关的个体与集体

1. 论文著者

著者是论文的生产者，并有其鲜明的标志，极易计数。就一般而言，著者可分为四种类型：个人著者，二人以上的合著者，团体机构作者，匿名作者。《科学论文著者人名录》是进行著者分析的

重要来源资料。它罗列有《近期目次》(Current Contents)所报道的全部论文主要著者的姓名及其地址。各种检索工具累积索引的作者部分,也是测量著者的一个重要工具。许多检索工具都对二人以上著者,多重揭示,有的互设参见,也可以解决多著者的统计。团体机构作者在文献计量学中,作为研究对象并不多,对其在相应的检索工具中进行统计也不困难。不管是个体还是集体著者,由于他们都与文献发生关系,就必然成为文献计量学的重要研究对象。

2. 读者

读者即文献的使用者,也是文献的社会效益与物质效益的体现者。图书情报部门对读者的情况是可以掌握的。既可以发证来作为依据,也可从阅读现场登记来统计。读者的层次不等,水平不一,所以对读者应加以区别对待。从读者那里收回反馈信息,特别是对文献的需求与评价、出版物的选择等,都是至关重要的。探索读者的意向、要求和使用文献的习惯,可以用来改善文献服务状况。要使文献计量学扩展其实用价值,必须围绕读者来开展工作,使文献之“矢”,对准口径而又高效率地射中读者之“的”。

(二) 出版物

出版物对文献计量学是充分而且必要的条件。主要研究对象是:

1. 一次引文

现代科学论文绝大部分都附有“参考文献”或“引用文献”。这种“参考文献”或“引用文献”是论文作者为了交待、揭示若干背景情况而提供的。它们与正文有着密不可分的联系,因为是一次文献所引证的,常常称之为一次引文。一次引文的出现有多方面的原因,正像可以人体分泌物来检查体格那样,我们可以从一次引文来洞察科学论文的某些状况和特征。因此一次引文目前已成为对一次文献研究的重要入口。1963年创办的《科学引文索引》目前已成为一次引文数据的权威性来源。随着《科学引文索引》的不断完善与发展,一次引文数据的获得会变得更加为方便和准确,这也是对文

献计量学的一种有利支持。

2. 二次引文

为了区别一次引文，凡是被目录、文摘、文献报告等二次文献所摘储的一次文献，我们常称之为二次引文。一次文献的计数，由于必须将一次文献的物理实体找到，常常是很难做到的。这样就只能用一次文献的替代物——二次引文来代替。从而人们可以在二次文献里抽样，以统计出一次文献某些有价值的信息。

3. 三次文献

年鉴、手册、百科全书、书目之书目等三次文献，由于它们常常登载有业已统计整理好的数据，所以它们也成为文献计量学的辅助工具，从中可以捷足先登地选择必要的数字，以开展文献计量学的分析研究工作。

4. 期刊

由于期刊有报道迅速、专指性较高、容量均匀、刊期固定、文章单元宜于统计等优点，现在几乎是文献计量学有关载体的唯一来源。报道期刊的目录是大量的，最权威的要算是：创刊于1932年的《乌利希国际期刊指南》（Ulrich's International Periodical Directory），从这类目录中可以获得大量有关期刊的资料。

（三）词汇

科学文献是由自然词汇（符号）科学排列而形成的；文献的贮存与检索又都是用人工语言——情报检索语言来进行的。词汇在情报活动中有如此重要的作用，必须成为文献计量学的研究对象。具有特征性的词汇是：关键词（keyword），叙词（descriptor），导引词（catchword），限定词（identifiers）等。其中最重要的是叙词。

叙词（在我国长期被称为主题词）是由情报学家莫尔斯（C. Mooers）首先提倡和使用的术语。叙词所具有的意义没有必要与辞典中的解释相一致，词汇数量被严格地控制起来，它被认为是为